

2차원 바코드와 UCC/EAN-128을 이용한 생물자원 자동인식시스템

주 민 석[†] · 류 근 호^{**} · 김 준 우^{***} · 김 흥 태^{***} · 한 복 기^{****}

요 약

컴퓨팅 환경이 발전함에 따라 다양한 물리적 객체와 디지털 정보를 연동하는 자동인식 연구가 활발히 진행 중이다. 이러한 자동인식시스템은 다양한 산업분야에서 활용되고 있음에도 불구하고 보건 의료와 관련한 자동인식 기술의 접목은 아직까지 다른 산업기술 전반에 미치지 못하고 있는 실정이다. 이에 따라 의료장비, 혈액, 인체조직 등 보건 의료 용품의 자동인식에 관한 여러 연구가 진행 중이다. 이 논문은 인간 유전체 연구의 필수 연구재료인 생물자원을 대상으로 자동인식 기술의 적용 방안을 제안한다. 먼저, 자동인식기술 도입을 위해 사용 환경상의 고려사항을 정의하고, 조사과정 또는 실험을 통하여 적합한 형태의 태그 인터페이스로서 바코드를 선택하였다. 바코드 심볼로지는 2차원 바코드 심볼로지인 Data Matrix를 사용하고, 데이터 스키마는 국제적 범용성 추구를 위하여 UCC/EAN-128 기반으로 설계하였다. 제안된 기술들이 실제 환경에 적용되는지를 보이기 위한 어플리케이션을 개발하고, 이에 대한 실험 및 평가를 다음의 방법으로 수행하였다. 생물자원이 실제 보존되는 영하 196℃, 영하 75℃의 초저온 보존환경에서 바코드 인식실험을 한 결과 1.6초 내외의 평균 인식시간을 보이며, 데이터 스키마는 생물자원 활용 분야의 요구사항을 만족하는 것으로 평가되었다. 따라서 제안한 방법으로 생물자원의 정보처리 과정에서 정확성과 데이터 입력의 신속성이 제공될 수 있다.

키워드 : 생물자원, 바코드, 데이터 이동 매체, 자동식별

An Automatic Identification System of Biological Resources based on 2D Barcode and UCC/EAN-128

Minseok Chu[†] · Keun Ho Ryu^{**} · Jun-Woo Kim^{***} · Kim Hung Tae^{***} · Bokghee Han^{****}

ABSTRACT

As rapid development of computing environment, field of automatic identification research which interoperates with various physical objects and digital information is making active progress. Although the automatic identification system is widely used in various industries, application of automatic identification system in the field of medical health doesn't reach other industry. Therefore research in medical health supplies such as medical equipment, blood, human tissues and etc is on progress. This paper suggests the application of automatic identification technology for biological resources which is core research material in human genome research. First of all, user environment requirements for the introduction of automatic identification technology are defined and through the experiments and research, barcode is selected as a suitable tag interface. Data Matrix which is 2D barcode symbology is chosen and data schema is designed based on UCC/EAN-128 for international defecto standard. To show applicability of proposed method when applied to actual environment, we developed, tested and evaluated application as following methods. Experiments of barcode read time at 196 and 75 below zero which is actual temperature where biological resources are preserved resulted read speed of average of 1.6 second and the data schema satisfies requirements for the biological resources application. Therefore suggested method can provide data reliability as well as rapid input of data in biological resources information processing.

Keywords : Biological Resources, Bar Code, Data Carrier, Automatic Identification

1. 서 론

분자생물학(molecular biology)의 발전과 인간게놈프로젝트(HGP)의 성공으로 인해 의학 분야에서는 유전체 시료(DNA), 혈청, 혈장 등의 생물자원을 활용한 연구가 급속도로 증가하고 있다. 이러한 인간 유전체 연구 증가 흐름에 필수 요소인 생물자원은 역학·임상·진단 정보들과 면밀히 연계되어

[†] 정 회 원 : 질병관리본부 국립보건연구원 기술연구원
^{**} 중 심 회 원 : 충북대학교 전기전자 및 컴퓨터공학부 교수
^{***} 정 회 원 : 질병관리본부 국립보건연구원 보건연구사
^{****} 정 회 원 : 질병관리본부 국립보건연구원 보건연구관
논문접수 : 2007년 12월 7일
수정일 : 1차 2008년 5월 22일
심사완료 : 2008년 6월 14일

실험연구와 분석에 사용되고 있다. 이러한 연구·분석 과정에서 개별 생물자원 정보의 정확성에 대한 중요도는 매우 높다. 예를 들어, 생물자원의 수집 및 관리 과정에서 검체 보관용기에 수작업으로 식별자를 표기하거나 관련 정보를 키보드를 통해 컴퓨터에 입력함으로써 인하여 식별자를 잘못 입력하는 오류를 범하게 되었을 때, 최종적으로 생물자원을 활용하는 연구자는 원하는 연구 성과를 얻지 못하거나 연구 결과에 근본적인 오류를 내포하게 되는 문제를 초래하게 되기 때문이다. 또한 복잡한 구조를 가진 다량의 정보를 디지털화하기 위한 입력방식으로 키보드를 사용하게 되면 입력 속도가 느리며 사용자의 많은 노력을 요구하게 되므로, 결과적으로 물리적 객체의 정보들을 수작업으로 디지털화하는 것은 많은 처리 시간과 오류를 발생시킨다는 단점이 있다[1]. 이러한 단점을 극복하기 위해서 바코드, RFID 등의 태그 인터페이스를 이용한 자동인식(automatic identification)기술이 사용된다. 자동인식은 수기 인식으로 이루어지던 방법을 개선하여, 생물자원을 보존하는 환경에서 정확성과 신속성을 높여줄 수 있는 중요한 기술이다.

보건의료분야 연구에서 활용되는 의료장비나 의약품, 실험도구에서부터 혈액, 장기 및 조직(organs or tissues) 등은 관리되어지는 물리적 객체의 자동인식을 위하여 가장 대표적 태그 인터페이스인 바코드를 활용하여 정보처리를 함으로써 정확도를 높일 수 있었다. 그러나 이와 같은 의료용품이나 연구 대상물에 비해 생물자원에 대한 대규모의 관리는 비교적 짧은 역사를 가지고 있어 IT분야를 비롯한 다양한 기술이 요구되는 현실이다. 2003년 인간게놈지도(map of the human genome)가 밝혀지고, 생물자원의 분석을 통하여 개인의 유전적인 소인에 바탕을 둔 질병의 위험도에 대한 예측이 가능하게 됨에 따라, 최근에 들어 생애주기나 생활 환경별 예측을 위한 다수의 프로젝트에 의해 생물자원들이 수집되고 있다[2]. 아직까지는 생물자원 관리의 경우에 수기로 자원정보를 처리하거나 1차원 바코드를 식별 매체로 활용하여 기관 자체의 식별체계 또는 단순 일련번호 체계의 사용을 통해 제한적 정보 공유가 진행되고 있다. 이러한 생물자원 식별 방법은 기관 간 자원 정보의 교환 및 공유가 어렵고, 바코드 자체에 소량의 정보를 담고 있어 생물자원의 대량 생산에 부적합한 형태를 갖게 되어, 전산시스템을 위한 다양한 정보를 제공해 줄 수 없는 문제점을 안고 있다. 따라서 우리는 이 문제를 해결하기 위하여 생물자원 고유의 특성에 적합한 자동인식 시스템을 연구하였다.

생물자원 컴퓨팅 환경에서 정확성과 신속성을 높이기 위한 자동인식 기술의 적용 방안으로, 우리는 2차원 바코드와 UCC/EAN-128을 이용한 생물자원 자동인식시스템을 제안한다. 논문을 효율적으로 전개하기 위하여, 2장에서는 생물자원 자동인식기술의 연구배경과 요구사항을 다루고, 3장에서는 생물자원의 영하 196°C의 검체 보관환경과 보관용기의 특수성에 적합한 태그 인터페이스를 고찰한 결과, 바코드가 가장 유용함을 제시한다. 4장에서는 바코드 도입을 위해 고려해야 할 제반사항을 정의하고, 그에 따라 바코드 심볼로지의 선택, 데이터 스키마의 정의를 한다. 그리고 5장에서는

제안된 내용을 기반으로 하는 시스템 구현을 기술한다. 6장에서는 구현된 생물자원 자동인식시스템을 이용하여 실험 및 평가를 한다. 7장에서는 연구결과를 간략히 요약하고, 추후 연구사항에 대하여 다룬다.

2. 연구 배경

2.1 생물자원 자동인식기술의 필요성

생물자원이라 함은 조직, 세포, DNA 등의 실물과 이와 관련된 모든 정보를 망라하는 것으로서 생명과학 분야의 연구발전에 절대적으로 필요한 재료이다[3]. 생물자원의 관리 체계화가 중요한 이슈로 대두되면서 자원의 수집과 보관과정에서 정보처리학, 산업공학 등의 분야와 연구기반 접목을 필요로 하고 있다. 그러나, 생물자원은 보관용기와 보관환경이 유통물류나 산업기반시설에서 활용되고 있는 양상과는 다른 면모를 보여 자동인식기술의 도입을 이루기가 쉽지 않다. 보관용기는 RFID 태그나 바코드를 부착하기에 협소한 표면적을 갖고, 상온에서부터 영하 196°C까지의 온도 변화에 견딜 수 있어야 하는 보관 환경적 특징을 갖기 때문이다. 따라서 보건의료분야 혹은 비슷한 환경에서의 객체식별에 관한 연구를 살펴보고 적용가능성을 검토하는 등의 관련 분야의 다양한 연구가 요구되고 있다.

보건의료분야에서 다루어지는 물리적 객체를 디지털 정보와 유기적으로 연계하는 연구는 크게 두 가지로 분류되어진다. 첫 번째로 정보처리의 관점에서 식별하고자 하는 제품이나 생물자원 등을 물리적 객체로 정의하여, 컴퓨팅 환경에서 식별할 수 있도록 식별 코드체계를 정립하고 자동인식 기술을 도입하는 영역이 있다. 두 번째로 생물학적 관점에서 모든 생물체는 제각기 다른 염기서열을 갖기 때문에 유전체(genome) 내에 존재하는 유전변이(genetic variation)를 확인함으로써, 개체(individual) 또는 종(species)의 차이를 식별하여 유일 식별자로 디지털 정보화를 하는 영역이다. 본 장의 2절에서는 바코드 시스템을 도입해서 활용하고 있는 보건의료분야의 바코드 시스템 적용사례를 살펴보고, 3절에서는 대표적인 생물학적 관점의 객체 식별방식인 DNA Barcode에 관하여 살펴본다.

2.2 보건의료분야의 바코드 적용사례

보건의료분야에서 사용되는 바코드도 다른 산업분야와 마찬가지로 식별하고자 하는 대상객체의 특성에 따라, 바코드 심볼로지나 데이터 표준을 정하여 사용하고 있다. 바코드를 사용하는 주요 보건의료분야는 의료장비, 의약품, 실험기구 등의 물품 객체와 수혈용 혈액, 장기 및 조직, 혈관계 조직 등의 생물로부터 파생된 객체가 있다.

이들 대상 객체를 식별하기 위한 바코드 심볼로지의 결정은 유통에 관여된 산업체 또는 관계기관에 의해 결정이 되는데 일반적으로 국제 바코드 표준기구(GS1)에서 제안하는 EAN-13이나 UCC/EAN-128을 응용하였다. EAN-13은 국제 거래단위번호(GTIN)로 식별정보를 표현하며, UCC/EAN-128

은 GTIN을 포함한 제품의 이동, 추적, 보관, 생산관리 등에서 요구되는 다양한 정보를 AI(Application Identifier)를 추가 사용하여 정보를 표현하기 때문에 식별정보 이외의 부가 정보를 담을 수 있는 특징이 있다[4]. GS1에서 제안되지 않은 형태의 바코드 심볼코지를 사용하는 바코드화된 객체도 있다. 혈액의 경우 국제 혈액은행 자동화 협회(ICCBBA)에서 제안한 ISBT128 바코드[5]를 사용하며, 미 보건의료분야에서는 GS1 바코드와 병행하여 보건산업협회의기구(HIBCC)에서 제안한 HIBC 바코드[6]를 사용하여 의료장비, 혈관계 조직을 식별한다. ISBT128은 Code128[7]을 보완한 코드체계로 6가지의 식별코드를 조합해 혈액제공자 식별자(donation ID), 혈액 그룹(blood groups), 유통 기한(expiry date) 등 혈액 및 혈액 구성요소를 식별한다. ISBT128은 혈액유통 분야에서 필요로 하는 고유의 식별정보와 추가적인 정보를 동시에 갖도록 하여 정보 시스템간의 직접적 연계가 없이 바코드만으로 최소한의 데이터 교환이 가능하도록 설계하였다. 그리고 HIBC 바코드는 ‘Where’의 개념으로 어디서 사용되는지를 정의하고 ‘What’의 개념으로 어떤 객체인지를 정의함으로써 의약품이나 의료기구에 대한 식별 뿐 아니라 보건의료분야에서 범용적으로 사용할 수 있게 설계하였다. <표 1>은 바코드가 활용되고 있는 보건의료분야의 실태를 GS1에서 조사한 내용이다[8].

2.3 DNA Barcode

DNA의 구성물질 중 염기는 아데닌(A), 티민(T), 구아닌(G), 시토신(C)의 네 종류로 나뉘어져 있다. 그리고 아데닌은 티민과, 구아닌은 시토신하고만 결합하는데, 이러한 결합을 염기쌍이라 한다. 이러한 염기쌍은 아메바나 박테리아

<표 1> 국제 보건의료산업 바코드 사용현황

분야	포장 단위	바코드 사용현황
의료장비 (Medical Devices)	단품 포장	· 2%: EAN13 · 98%: 바코드 없음
	박스 포장	· 약60%: EAN13+ 128 · 약30%: EAN128 *미국: HIBC *일본: JFMDA(EAN128)
	운송용 포장	· 약60%: EAN13 + 128 · 약30%: EAN128 *미국: HIBC *일본: JFMDA(EAN128)
의약품 (Pharmaceuticals)	단품 포장	· 250ml미만 제품: 바코드 없음 · 250ml이상 제품: 약90% EAN13 *미국: RSS(NDC, batch no, expiry)
	운송용 포장	· 약90%: EAN13
실험도구 (Instruments)	단품 포장 박스/운송용 포장	· 100%: EAN128
혈액유통 (Blood)	단품 포장 박스/운송용 포장	· 100%: ISBT128
장기 및 조직 (Surgical Products)	박스 포장 운송용 포장	· 약90%: EAN128
혈관계 조직 (Vascular Systems)	박스 포장 박스/운송용 포장	· 95%: EAN128 *미국: HIBC 병행사용



(그림 1) 유전자신분증제도 데이터 시스템

등 단세포 생물에서부터 인간에 이르기까지 모든 생물에서 공통적으로 볼 수 있는 현상이며, 모든 생물체는 A·T·G·C 네 종류의 염기의 배열방법이 유전정보 그 자체이면서 생물의 종류와 종에 따라 제각기 다른 배열방법을 갖는 특징이 있다. DNA Barcode는 미토콘드리아 유전체에서 Cytochrome C Oxidase I(COI)으로 알려진 648개 염기쌍 영역의 배열이 각기 생물에 따라 다르고, 생물종에 따라서 다르며, 중복되지 않으므로 종분류 식별자로 사용할 수 있다고 제안하였다[9, 10]. 이 식별자를 표현하기 위하여 2차원 바코드를 매체로 사용하였으며 A·T·G·C 값을 이용하여 4진수의 연속된 정보를 담아 정보처리가 원활하게 하는 특징을 갖고 있다.

최초 논문 발표이후, 제안된 내용을 기반으로 하는 각종 동·식물에 대한 실험연구가 진행되었다. DNA Barcode를 활용하여 새에 대한 종분류 실험연구가 발표되었고, 개화 식물에 대한 종분류, 고대 생물에 대한 종분류 등의 연구가 진행되었다. 현재 실험연구를 통해 검증된 DNA Barcode의 유용성이 인정되어 국제적으로 생물의 종판별 표준시스템이 정착되어 있으며, 유전자신분증제도(barcode of life)를 실용화하기 위하여 (그림 1)과 같이 생물자원 식별정보를 DNA Barcode화하여 등록·저장하는 시스템을 웹을 통해 일반인에게 공개하고 있다.

2.4 생물자원 자동인식을 위한 요구사항

기존의 적용사례와 이 논문의 대상 객체인 생물자원은 생명과학분야라는 점에 있어서 유사하다. 그러나 식별하고자 하는 대상 객체의 특성과 활용 용도에 따라 적용되는 기술이 다양함을 2.3절의 사례를 통해 알 수 있고, 더욱이 보건의료분야 인체유래 생물자원을 대상으로 하는 바코드 표준체계가 발표된 사례가 없는 것으로 조사되어 다양한 관련 연구가 요구되고 있다. 또한 DNA 바코드와 같은 생물학적 관점의 객체 식별방식은 종분류까지만 식별이 가능하여 개별 객체를 식별하지 못하므로 보건의료분야 생물자원의 보존관리에서 물리적 객체 식별자로 적절하지 않다.

최근들어 생물자원을 대량으로 수집·보관하는 일의 중요성이 대두되면서 자원은행의 개념이 도입되면서 유통이나 물류에서 활용되었던 대용량 객체정보처리의 개념을 요구하고 있다. 이에 따라, 대량의 객체정보처리에서 오류를 최소화하고 데이터 입력방식 개선을 통한 신속성 향상을 가져올 수 있는 자동인식기술의 도입연구가 보건의료분야 자원은행

분야에서 요구되고 있다. 이와 함께 표준화된 생물자원 식별자 표현스키마 부제로 인한 정보교류의 혼란을 해결하는 방법이 고려되어야 한다.

3. 태그 인터페이스 분석

바코드, RFID, 스마트 카드 등의 태그 인터페이스 도입시 생물자원 관리 환경에서의 고려사항 정의가 선행되어야 한다. 이 논문은 태그 인터페이스 도입을 위한 고려사항을 초저온 환경, 내습성, 태그소형화 가능성, 도입비용으로 정의한다. 생물자원은 검체 종류에 따라 액화질소탱크(-196℃) 또는 초저온 냉동고(-75℃)의 보존환경을 요구하므로, 상온에서부터 영하 196℃까지의 온도 변화에 견디고 태그 인터페이스에 표현된 정보가 잘 보존되는지의 초저온 환경을 고려사항 중 하나로 본다. 그리고 검체는 마이크로리터(ul) 단위의 부피량으로 연구에 활용되므로, 실제 보존·유통하는 부피가 밀리리터(ml)단위의 작은 용기에 보관되고 있기 때문에 인식정보를 담기에 매우 작은 표면적을 갖는다. 따라서 자원 보관용기를 고려하여 10×10mm 크기의 태그(tag) 부착 표면적으로 정의되는 태그소형화 가능성을 고려한다. 도입비용은 생물자원을 수집·보관하는 생물자원은행이 보유하고 있는 자원의 개수를 200만개라 가정하고, 객체에 태그 인터페이스를 도입하였을 때의 비용을 부착 태그당 가격으로 산정해 고려사항의 요소로 정의한다.

정의된 고려사항에 대해 현존하는 태그 인터페이스별로 도입 가능성에 대해 분석하고, 적합한지의 여부를 나열하면 <표 2>와 같다. ×는 부적합함을 나타내고, ○는 적합함을 나타낸다.

초저온 환경과 내습성은 전자태그 기반의 RFID나 스마트 태그의 경우 극저온에서는 부품 소자의 수축 현상으로 제품 파손이 일어나 데이터가 유실되는 반면, 바코드는 저온 환경에서 견딜 수 있는 기술이 존재한다. 특히 바코드 기술은 바코드를 인쇄하여 부착하는 방식과 용기 표면에 잉크젯 마커나 레이저 마커로 보관용기에 직접적 표기를 하는 방법 등 다양한 형태의 표현방식이 존재하므로 생물자원을 관리하는 기관이나 환경에 따라 여러 가지 방식을 선택할 수 있는 장점을 갖는다.

태그소형화 가능성은 전자태그기반 태그 인터페이스의 경

우 현재 다양한 연구가 진행되고 있으나 제시된 10×10mm의 크기에 적합한 산업기술이 존재하지 않으며, 현재 Smart Dust 프로젝트[11] 등을 통하여 전자태그의 소형화에 대한 연구가 진행된 바 있으나 태그소형화와 함께 초저온 환경과 내습성에 대한 동시 고려되지 않아 적용가능한 수준이라 보기 어렵다.

도입비용은 대한무역투자진흥공사(KOTRA)에서 발표한 ‘2006년 미국 지역 스마트카드 및 리더기 시장동향’을 기준으로 전자태그기반 태그 인터페이스 가격을 산정하고, 바코드의 경우 시장조사를 통해 내한용 특수라벨용지, 바코드리본의 가격을 더하여 산정하였다. 산정결과 200만개의 객체에 도입하는데 드는 도입비용이 RFID > Smart Tag > Barcode 순으로 나타났다. RFID의 경우 바코드시스템의 무려 60배에 가까운 것으로 조사되었다.

태그 인터페이스를 고찰한 결과 생물자원의 자동인식기술 도입을 가능하게 하여 자원관리 효율성을 가져다 줄 기술은 바코드 기술이라 할 수 있다. 최근의 IT 분야에서 유비쿼터스 시대를 선도할 기술로 RFID, 스마트 카드 등의 매체를 뽑고 있지만 전자태그 기반 기술의 이면에는 상황에 따라 달라질 수 있는 태그 인식률, 사생활 침해나 보안과 관련된 논란, 실용화하기에는 고가의 태그 가격 등 해결해야 할 과제가 산적해있고 인간의 건강과 생활을 다루는 생물자원에 접목하기에 아직까지는 기술 검증이 충분치 않은 상태이기 때문이다. 그에 비해 바코드 기술은 끊임없이 다양한 산업 분야에서 도입되면서 다양한 형태로의 연구가 진행되고 그 기술이 검증된 상태이다.

최근 각광받는 RFID도 기술의 오작동이나 파손을 보완하기 위한 최적의 솔루션이 바코드 기술이라고 평가받기도 한다. 예를 들어 기존의 바코드 라벨 뒷면에 RFID 태그가 내장된 형태의 스마트 라벨(smart label)의 연구는 특정 물건에 RFID 태그를 부착한 후 리더를 통해 데이터 처리를 아무리 잘 해낸다 하더라도 기하학적인 패턴만을 지니고 있는 RFID 태그를 사람이 육안으로 판별할 수 없기 때문에 그 보완 기능이 동시에 사용되어야 한다는 점을 착안하여 사용되고 있으며, RFID의 오작동을 보완하는 기술인 RFID/2D 혼합코드 연구도 바코드 기술의 장점 요소를 부각시키는 맥락에 있다. 유비쿼터스 시대에 전자태그 기반의 기술이 보편화 된다고 하여도 바코드는 구시대적 유물이 아닌 다른 기반기술을 보완해주는 요소로 사용될 것이고, 적어도 생물자원 관리환경에서는 가장 적합한 태그 인터페이스로 볼 수 있다.

<표 2> 태그 인터페이스의 생물자원적합성 비교 결과표

고려사항	전자태그기반		인쇄기술기반
	RFID	Smart Tag	Barcode
초저온 환경 (50℃ ~ -196℃)	×	×	○
내습성	×	×	○
태그소형화 가능성 (10×10mm)	×	×	○
도입비용 (총액/단가)	60억원 / 약 3,000원	20억원 / 약 1,000원	1억원 / 약 50원

4. 생물자원 자동인식시스템

4.1 바코드 심볼로지 선택

바코드 매체를 활용하기 위해서는 디지털화하기 위한 대상 객체를 정의해야 한다. 생물자원 관리환경에서는 이들 자원의 제공자를 식별할 수 있는 정보와 자원의 종류가 중요한 정보로 활용되며, 개개의 보관용기도 의미있는 정보로서 유일하게 식별되어야 한다. 따라서 생물자원 관리환경에서 물리적 객체란, 생물자원을 담는 자원 보관용기와 박스(box), 랙(rack) 등의 운송·보관용 포장용기라고 정의할 수 있으며, 자원 보관용기는 담겨진 생물자원에 대한 추가 정보를 제공할 수 있어야 한다.

운송·보관용 포장은 그 크기나 활용 용도가 일반 유통의 목적과 다르지 않지만, 자원 보관용기는 바코드를 인쇄하기에는 협소한 표현면적을 갖고 있고, 원통형 구조로 되어있으며, 용기에 담긴 생물자원을 식별해야 하는 제약조건이 있다. 자원 보관용기도 일반 유통의 목적을 포함해야 하므로 운송·보관용 포장은 자원 보관용기의 바코드 심볼로지에 필요조건을 만족한다. 따라서 운송·보관용 포장은 자원 보관용기의 분석에 따라 정해지는 바코드 심볼로지를 동일하게 적용할 수 있으므로, 자원 보관용기의 특징에 적합한 바코드 심볼로지를 살펴본다. 이 논문은 생물자원을 보관하는 가장 보편화된 용기인 e-tube(1.5ml eppendorf tube)와 냉동용기(1.8ml cryo tube)를 자원 보관용기로 정의한다.

자원 보관용기에 라벨링이 가능한 너비(원둘레)와 높이는 (그림 2(a))의 e-tube의 경우 약 35 × 17mm이며 (그림 2(b))의 냉동용기는 약 40 × 32mm 이므로, 동일한 바코드 심볼로지를 사용하였을 때, 최대 표현면적은 35 × 17mm 이다. 보관용기는 (그림 2)와 같이 라벨링을 하는데 있어서 크기가 매우 작은 편이고 원통형 구조를 가지고 있어 인쇄된 바코드를 읽을 때 굴곡에 의한 데이터 왜곡현상이 발생한다. 또한 생물자원은 연구 활용 이전까지 반영구적으로 보존을 하며, 특히 동결 보존한 이후에는 표현된 라벨을 변경할 수 없는 특징을 가지고 있다. 왜냐하면, 동결보존 중에 있는 세포를 해동(thawing)시 적절한 절차를 따르지 않는 경우 세포외액으로부터 세포내로 수분이 유입되면서 세포내의 얼음 재결정화(recrystallization)현상이 발생하게 되어 세포에 치명적인 손상을 입히게 되기 때문이다. 따라서 바코드가 라벨링이 되고 동결보존이 되면 라벨링 변경은 사실상 불가능하기 때문에 지속적 기술발전에서 기존의 기술이 고려되

어질 수 있도록 바코드 심볼로지의 국제적 범용성을 갖춘 바코드 심볼로지를 사용해야 한다.

보관용기의 최대 표현면적과 바코드 심볼로지의 국제적 범용성을 고려하여 현존하는 약 150여개의 바코드 심볼로지 중에서 EAN-8, UCC/EAN-128, PDF-417[12], QR Code[13], Data Matrix[14]를 후보군으로 선정한다. 선정기준은 다음과 같이 정하며, 선정된 5가지의 바코드 심볼로지에 대한 적용 가능여부 실험 및 평가한다.

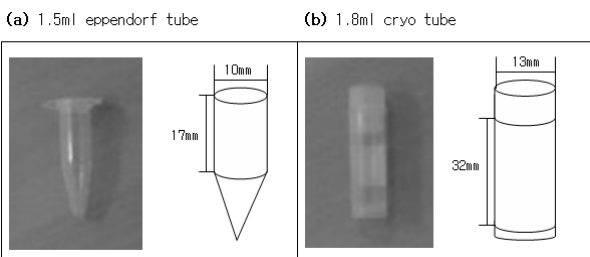
- 35×17mm 면적에 표현 가능한 바코드 심볼로지. 단, 바코드 리더기가 코드 인식을 위한 여백과 가독문자를 표현할 수 있는 최소공간이 확보되어야 한다.
- ISO 국제표준에 의해 관리되며, 특정 산업기반에 한정되지 않은 바코드 심볼로지

4.2 데이터 스키마 정의

1차원 바코드 심볼로지가 심볼로지에 국한된 데이터 스키마만을 사용하는 반면에 Data Matrix, QR Code 등의 2차원 바코드 심볼로지는 심볼로지 데이터 영역(data region)에 ASCII-128을 기본적으로 담을 수 있게 고안되어있다. 따라서 2차원 바코드는 원하는 형태의 데이터 스키마에 따라 정보를 입력할 수 있다. 그러나 생물자원 식별을 위한 네이밍스페이스(naming space)나 분류체계에 대한 국제표준이 존재하지 않기 때문에, 이 논문에서는 데이터 스키마 정의를 위한 요구사항을 정리하고, 보편적으로 사용되는 UCC/EAN-128 기반의 데이터 스키마를 정의한다.

연구배경에서 기술한 생물학적 관점의 개체식별방법인 DNA 바코드는 국제적으로 통용되는 바코드 표준에 기반한 데이터 스키마를 갖지 않는다. 또한, 바코드의 활용용도가 일반적인 정보처리 과정에서 정확성과 데이터 입력의 신속성을 높이기 위함이 아닌 동·식물에 대한 종 분류 정보를 648개 염기쌍 영역의 배열로 표현하는 하나의 기법으로 바코드를 사용한 경우라고 볼 수 있다. 반면, 본 논문에서는 데이터 스키마를 검체보관용기를 유일하게 식별하기 위한 정보만을 사용하였기 때문에 국제표준 바코드를 사용할 수 있게 된다. 그 이외의 용기 내에 담겨진 내용물에 대한 정보는 데이터베이스로 관리하는 형태를 지향한다. 바코드와 연계되는 데이터베이스는 관리되는 유형에 따라 일반 동·식물의 경우 개별 종의 분류학적인 의미 등이 중요시 관리되고, 인간의 경우 일반 동식물과는 달리 질병과 관련된 유전자의 발견이나 신원 확인, 친자 확인 등 검체제공자 개별 존재에 대한 의미가 중요시 관리되므로 휴먼게놈프로젝트에 의해서 밝혀진 인간의 30억 쌍의 염기서열을 범의학에서 주로 사용하는 STR기법을 이용한 DNA 분석데이터나, SNP 분석결과 등을 개별 데이터베이스로 관리하는 형태를 지향한다. 따라서 검체를 관리하기 위한 직접적인 데이터를 제외한 정보는 데이터 스키마의 요구사항에서 제외하였다.

데이터 스키마 요구사항을 분석하기 위하여, 의학, 임상병리학, 생물학, 유전체학 등의 생물자원을 활용하는 집단의 전문가들로부터 자문을 얻어 데이터 스키마가 만족시켜야 할 요구사항을 발견·정제하고, 보편적으로 식별자가 갖는



(그림 2) 생물자원 보관용기

특징을 조합하여 다음과 같이 정리한다.

- 유일성(uniqueness): 식별코드는 오직 한 객체에 할당되어야 하고 재사용되어서는 안된다.
- 영·숫자(alphanumeric) : 식별에 사용되는 데이터 문자(data characters)는 영문자와 숫자가 표현되어야 한다.
- 국제성(international): 생물자원의 공급·유통에 관여되는 모든 집단들이 객체의 생명 주기가 존재하는 동안 동일한 식별코드와 식별 시스템을 공유할 수 있도록 국제표준에 의해 지정되어야 한다.
- 중립성(neutral) : 특정 응용프로그램에 종속되지 않고 다양한 응용프로그램이나 단체들의 공통된 식별코드로서 역할을 해야 한다.
- 영속성(persistent) : 식별코드가 객체를 식별하지 않더라도 수명이 제한되어서는 안된다.
- 유연성(flexible) : 식별코드는 객체를 사용하는 기관마다의 사용목적이나 식별기준이 서로 다를 수 있으므로 식별 정보로서의 요소를 유연하게 사용할 수 있어야 한다.
- 입도(granularity) : 식별 시스템은 시료를 담은 패키지가 단위 객체 단위로 구성된 것처럼 단일항목 또는 복합항목의 의미를 표현할 수 있어야 하고 구성요소들은 단일 객체로서 정의되거나 묘사되어야 한다.
- 용량(capacity) : 관리되어지는 생물자원의 양이 매우 많은 것에 대해 포용이 가능하여야 한다. 식별시스템은 생물자원의 보관용기가 작은 특수성을 지녔어도 시스템의 범위 내에 정의된 객체의 용량을 위해 충분한 데이터표현 양을 지녀야 한다.
- 부착면적(area) : 식별시스템이 식별하고자하는 생물자원이 매우 작은 보관용기에 담기는 것을 감안하여 정의된 데이터 스키마가 생물자원 보관용기 면적에 표현이 가능한 심볼로지 기술이 존재하여야 한다.

데이터 스키마 요구사항을 만족시키기 위하여, UCC/EAN-128을 기반으로 데이터 스키마를 설계하였다. 대표적인 바코드 데이터 스키마 표준인 GS1 표준코드 중에서 EAN-8, EAN-13, EAN-14는 유통물류에서 단품 또는 박스단위에 인쇄되어 매우 활발히 활용되었으나 무의미성, 범용성 한계 등이 제기되었고, 이를 개선하고자 UCC/EAN-128 표준코드가 제정되었다. UCC/EAN-128 표준코드는 어느 국가나 장소, 또는 활용 산업분야에서도 사용가능한 개방형 국제표준이며, 관련된 산업분야에서 필요로 하는 새로운 정보가 발생할 경우, 데이터의 형식과 의미를 지정하는 프리픽스인 응용식별자(AI)를 새로 정의하여 사용할 수 있는 유연성을 가지고 있는 특징이 있다.

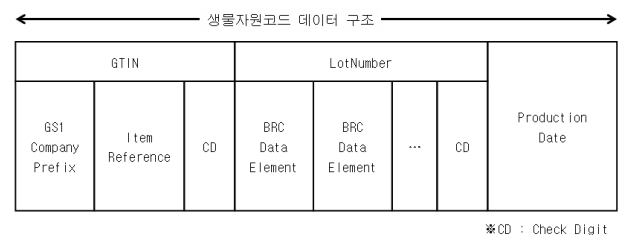
이 논문이 제안하는 데이터 스키마는 생물자원을 식별하는 코드이므로 “생물자원코드(BRC ; Biological Resources Code)”로 명명한다. BRC는 생물자원 식별 요구사항을 만족시키기 위하여 UCC/EAN-128 식별 시스템에서 제공하고 있는 국제거래단위번호(GTIN), 로트번호(lot number), 그리고 생산일자(production date)로 구성된 3가지 타입의 식별코드를 사용하여 개개의 생물자원 정보를 식별한다.

GTIN은 상품의 정보 검색과 주문이나 계약에 필요한 상품 및 서비스 단위를 유일하게 표기하기 위한 세계적 표준이다. 상품, 서비스, 거래처를 나타내는 코드는 국가, 제조업체, 품목별로 고유번호가 부여되기 때문에 전 세계적으로 중복되는 코드가 없다. 따라서 한번 부여된 번호는 관련 산업계에 있는 모든 기관들이 동일한 식별번호로 활용할 수 있을 뿐만 아니라 부여된 번호를 중심으로 정확한 정보를 상호 교환할 수 있다. 따라서 국제성, 중립성을 보장한다. GTIN에 물류식별자(indicator)는 낱개 상품에서부터 각각의 포장단위(박스 등)에 대하여 1부터 차례로 숫자를 부여하는 규칙을 가지므로 입도를 보장한다. 또한 로트번호는 물류단위에 대해 중량, 사이즈 또는 길이 등이 일정하지 않은 상품인 경우 또는 주문자의 주문에 의해 생산되는 경우에 물류단위의 중량, 포장된 상품의 개수 등 세부적인 표현이 가능하다. 따라서 코드 유일성과 영속성, 유연성, 용량의 요구사항에 대한 만족을 위해 이 논문에서는 로트번호를 활용하여 세부 필드데이터 규칙을 정의한다. UCC/EAN-128 표준의 데이터 표현문자는 ASCII를 사용하므로 영·숫자의 요구사항이 만족된다. 또한, BRC는 전체 데이터 자리수가 30자 내·외로 구성되며 Data Matrix, QR Code의 바코드 심볼로 지로 표현이 가능한 기술이 존재하므로 부착면적에 대한 요구사항을 만족시켰다고 볼 수 있다.

BRC의 데이터 구조는 (그림 3)과 같이 GTIN, 로트번호, 생산일자를 데이터 요소로 구성하여 생물자원 정보를 나타낸다. GTIN은 공급망에서 상품의 정보 검색과 주문이나 계약에 필요한 상품 및 서비스 단위의 식별을 위해 사용되는 번호로서 거래 항목을 정의하기 위한 국제적인 표준이다. 이 데이터 요소는 생물자원의 수집기관, 관리기관 및 자원 활용 연구기관간의 유통과정에서 물리적 객체 식별을 만족시킬 수 있는 데이터 요소라고 볼 수 있다.

GTIN을 이용하여 기본적인 유통 환경을 위한 제품식별은 가능하지만 보건의료 및 생명공학분야에서 사용되는 생물자원은 단순한 제품식별정보만으로는 종종 불충분한 경우가 있다. 연구 목적으로 활용이 되는 생물자원은 자원의 생성에서 소멸에 이르기까지 자원이 담고 있는 세부내용이 더욱 중요하게 여겨지는 경우가 보편적이기 때문이다. 따라서 로트번호는 생물개체를 위한 개체식별자, 객체의 종류 그리고 객체번호를 <표 3>과 같이 필드화하여 구성함으로써 기관간의 정보교류시 상호 운영성을 높일 수 있다.

로트번호는 수식 (1)에서 표현된 것처럼, 하나 또는 그 이상의 생물자원 데이터요소(B)로 구성이 된다. 생물자원 데



(그림 3) 생물자원코드 데이터 구조

<표 3> 로트번호 필드데이터 구조

필드명	플래그 문자	ASCII	필드설명
개체식별자	#	35	생물개체(검체제공자)를 유일하게 식별하기 위한 데이터 필드
객체종류	\$	36	생물체로부터 파생되어 시험, 검사, 분석이 쓰이는 시료의 종류 (예, 세포, DNA, 혈청, 혈장 등)
객체번호	!	33	물리적 객체(보관용기)를 유일하게 식별하기 위한 데이터 필드

이더요소는 <표 3>의 필드데이터를 나타내는 요소들로 식별하고자하는 생물자원의 세부항목을 표현한다. 생물자원 데이터요소는 수식 (2)의 필드데이터(F)와 데이터요소(D)의 조합으로 구성이 된다. 따라서 로트번호는 수식(1)과 같이 정의할 수 있다.

Number of BRC Data Element := n

$$Lot\ number(n) = null \quad \text{if, } n=0$$

$$= B_1 + B_2 + \dots + B_n + CD \quad \text{if, } n>0 \quad (1)$$

$$B = F + D \quad (2)$$

로트번호에서 체크문자는 마지막 자릿수를 뺀 나머지 문자들이 전산 상에서 올바르게 부여되었는지 확인하는 기능을 한다. 체크문자(CD)의 생성은 제어 문자를 포함하여 사용된 바코드문자의 합을 구한 후에 94 modulo 방법에 의한 나머지 값으로 적용하며 수식 (3)에 의해 계산된다.

Number of Lot number Character := n

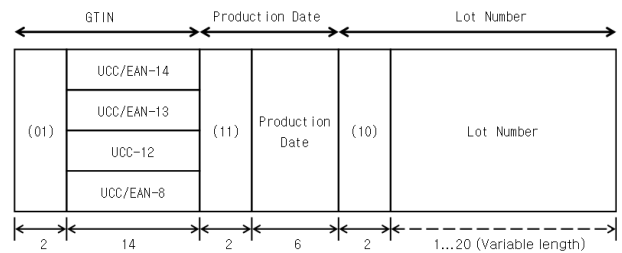
$$CD(n) = \left(\sum_{i=1}^n C_i \right) \text{mod } 94 \quad (3)$$

생물자원의 생산일자는 특정 시점을 대표하는 정보로 사용된다. 또한, 생물자원 보존·관리 과정의 생물 안전성에 대한 연구에서는 보존방법, 온도 등과 함께 보존 기간의 정보가 중요한 정보로 사용된다. 보존 기간의 경우는 생물자원 생산일자에서부터 연구 시점까지의 정보이며, 생산일자를 통해 정보 획득이 가능하다. 이와 같이 제품에 대한 생산일자를 나타내는데 사용되는 생산일자 코드는 년, 월, 일의 데이터를 표기한다.

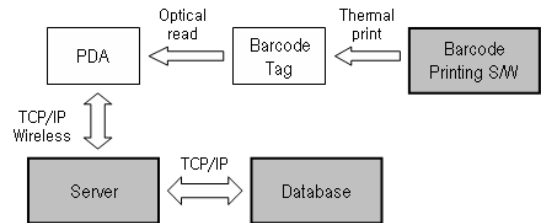
(그림 5)의 데이터 구조는 UCC/EAN-128 코드체계를 이용해서 (그림 4)과 같이 데이터 표현이 가능하다. 로트번호는 20자내의 가변형 데이터로, GTIN과 생산일자의 고정형 데이터를 파싱하고 남은 데이터는 가변형의 로트번호이기 때문에 데이터 파싱의 편의성을 줄 수 있다. 이와 같이 표현된 생물자원코드는 2차원 바코드 심볼로 생물자원 보관용기에 부착하여 면적에 대한 제약조건을 해결한다.

5. 시스템 구현

이 논문에서 제안한 생물자원 자동인식시스템이 어떻게 실제 환경에 적용되는지를 보이기 위하여 어플리케이션을 구현한다. (그림 5)은 전체시스템의 블록 다이어그램이다. 구현한 시스템은 바코드 프린팅 S/W, Server, 데이터베이스로 구성된다.



(그림 4) 생물자원코드 데이터 표현형



(그림 8) 전체시스템 블록 다이어그램

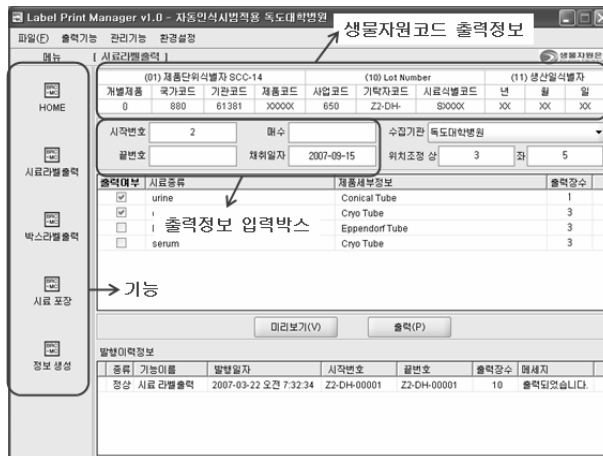
바코드 프린팅 S/W는 자바(java)기반으로 구현하고, Server는 JSP 프로그래밍 언어를 사용한 웹기반 응용프로그램으로 작성한다. 그리고 데이터베이스는 오라클을 사용하여 관계형 데이터모델로 구현한다. 바코드 프린팅 S/W는 감열형 바코드프린터기를 통하여 생물자원코드를 Data Matrix 바코드형태로 표현한다. PDA는 바코드를 읽을 수 있는 모듈을 탑재하면서 무선 네트워크로 서버에 정보를 주고 받을 수 있는 산업형 PDA를 채택한다. 서버에 개발된 웹 기반의 응용프로그램은 PDA에 데이터베이스에 수록된 정보를 보여주거나 관리할 수 있는 기능을 제공하므로, 바코드 태그를 읽어 정보처리가 가능하도록 하였다. 구현을 위한 개발환경은 H/W와 S/W 환경으로 구분하여 <표 4>에 기술한다.

바코드 프린팅 S/W는 식별자를 데이터 스키마에 맞게 입력하고 Data Matrix에 담아 바코드 프린터기로 정보를 출력할 수 있는 기능을 보관용기 및 포장용기별로 제공한다. 그리고 식별자의 구성요소에서 변경 가능성이 적은 국가코드, 기관코드, 제품코드 등의 정적데이터를 프로그램 환경설정을 통해 미리 지정이 가능하게 하였으며, 개별 객체에 따라 변경되는 생물자원제공자(기탁자코드), 생물자원식별코드(시료 식별코드) 등은 출력정보 입력박스에 사용자의 직접입력으로 출력할 수 있게 한다. (그림 6)은 바코드 프린팅 S/W의 보관용기용 바코드 라벨링을 위한 프로그램 수행 화면이다.

생물자원 관련 정보의 검색과 저장위치지정, 정도관리, 자

〈표 4〉 생물자원 자동인식시스템 개발 환경

구분		개발 환경	
H/W	Barcode Printing S/W	PC	Pentium 4 CPU 2.8GHz, Memory 1GB
		Printer	SATO CL412e, 300dpi Thermal Printer
		Reader	COGNEX DM7550
	Barcode Tag	Label	Barall bil-704, 유포지
		Ribbon	ARMOR AX8, Super Resin
		PDA	Symbol MC50, 산업형 PDA
Server	SUN V440, CPU 1.5GHz*2, Memory 4GB		
S/W	Barcode Printing S/W	JDK, Java applications	
	PDA	Pocket Internet Explorer, Web Browser	
	Server	JDK, JavaServer Pages	
	DBMS	Oracle 9i	



(그림 6) 바코드 프린팅 S/W 수행화면



(그림 7) PDA로 본 생물자원관리시스템

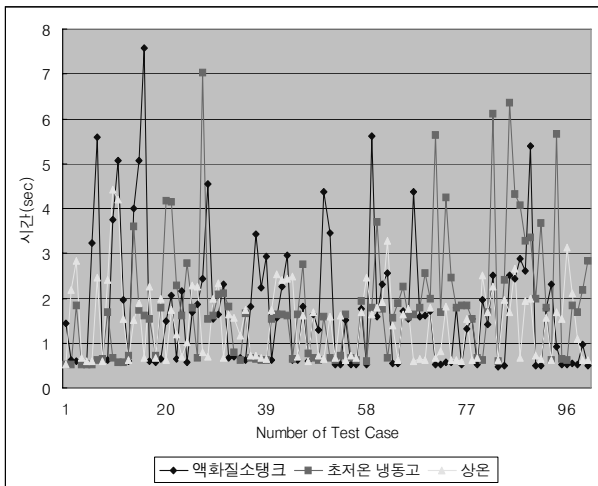
원 분주 등의 웹기반 관리서비스는 (그림 7)과 같이 웹 브라우저를 통해 이루어지고, 바코드 인식기능이 탑재된 PDA에서 무선 네트워크를 이용하여 정보를 처리한다.

6. 실험 및 평가

이 논문에서 제시하는 생물자원 자동인식시스템이 활용적 가치를 얻기 위해서는 바코드 심볼로지가 극한의 저온환경에서 데이터 손실없이 인식률이 높아야 하며, BRC 데이터 스키마에 대한 생물자원 보관용기와 바코드 심볼로지의 적합성을 고려하여야 한다. 따라서 제시된 데이터 스키마를 준수하는 데이터를 바코드 심볼로지에 표기하여 실제 생물자원 보존환경에서 인식률을 실험하여 얼마만큼의 인식률을 갖는지를 확인하고, 제시된 데이터 스키마에 대해 생물자원 보관용기의 최대 표현면적과 후보군 심볼로지의 요구 사항을 만족하는지에 대해 비교 평가를 한다.

6.1 저온환경에서의 데이터 인식률

인식률 실험을 위해 액화질소탱크, 초저온 냉동고, 상온으로 3가지 환경을 구성하였고, 바코드리더기를 이용하여 인식시간이 얼마나 걸리는지를 확인한다. 관측대상 a 는 액화질소탱크에 저장되었던 보관용기, 관측대상 b 는 초저온 냉



(그림 8) 보존상태별 바코드 인식시간 실험결과

<표 5> 보존상태별 바코드 인식시간 비교

보존상태	바코드 인식시간 (Sec)			
	Min.	Ave.	Max.	Std-Dev.
액화질소탱크 (a)	0.48	1.7352	7.57	1.4476
초저온 냉동고 (b)	0.5167	1.8437	7.03	1.4016
상온 (c)	0.5417	1.4378	4.43	0.8364

동고에 저장되었던 보관용기, 관측대상 c는 상온 상태의 보관용기라 할 때, 저장상태에서 꺼내는 시간과 성애, 얼음덩어리 등을 제거하는 시간을 제외한 인식시간 T는 다음의 식으로 정리할 수 있다.

$$T = (\text{decode time}) - (\text{trigger time of barcode reader})$$

$$T_a = \text{액화질소탱크 보관 중 보관용기 바코드 인식시간}$$

$$T_b = \text{초저온 냉동고 보관 중 보관용기 바코드 인식시간}$$

$$T_c = \text{상온 상태의 보관용기 바코드 인식시간}$$

(그림 8)은 각 관측대상 a, b, c 마다 데이터가 성공적으로 읽힐 때까지의 인식시간 T를 측정된 결과이다. (그림 8)에서 X축은 바코드리더로 보관용기에 부착된 바코드를 읽은 수를 나타낸다. 실험결과, <표 5>에서와 같이 Ta, Tb, Tc의 결과가 평균적으로 1.6초 내외의 인식시간을 보이고 있다.

인식시간 평균은 $T_c < T_a < T_b$ 를 갖고 최대값은 $T_c < T_b < T_a$ 를 갖고 있으므로 상온에서의 인식률이 냉동 보존상태에서 보다 빠른 인식률을 보인다. 그러나 냉동 보존상태였던 Ta와 Tb 사이에는 근소한 차이를 보이고 있고 Tc와는 다소 차이 있음을 보이고 있으며 표준 편차 값이 상온보다 냉동 보존상태가 크다. 냉동 보존상태의 보관용기를 읽을 때 성애, 얼음 덩어리를 제거한다는 가정을 하였음에도 상온의 상태보다 인식시간이 긴 이유는 매우 극저온의 환경에 있었던 생물자원의 내부 온도가 상온에서 올라감에 따라 바코드를 읽는 동작시간에 급속도로 이슬맺힘 현상이

발생하여 바코드 리더기로부터 들어오는 광원을 분산시키기 때문이다. 바코드리더로 목표물을 조준하는 순간의 기계판독과 이슬맺힘의 정도에 따라 7.57초(테스트케이스 16), 7.03초(테스트케이스 27) 등의 느린 인식결과가 발생할 수도 있음을 보인다. 그럼에도 전체적으로 바코드를 읽는데 걸리는 시간은 평균 1.6초 내외의 인식시간을 보여주고 있다.

일반적 환경에서 바코드 인식시간은 1초 내외인데 비해 냉동상태 보관용기의 인식시간은 인식 순간의 환경적 변수에 영향을 받을 수 있다. 하지만 제약조건을 갖지 않는 일반 환경과 달리, 냉동 보존되는 생물자원의 특수성에 따른 다양한 제약조건을 감안하였을 때, 제안된 바코드 심볼로지는 활용적 가치가 있다고 볼 수 있고, 이동물체 영상인식분야, 비전기술(vision technology), 이미지 보정기술 등의 적용을 통하여 일반적 바코드 인식률과 동일한 결과를 가져올 수 있다.

6.2 생물자원 관리환경에서의 데이터 스키마에 대한 심볼로지 적합성

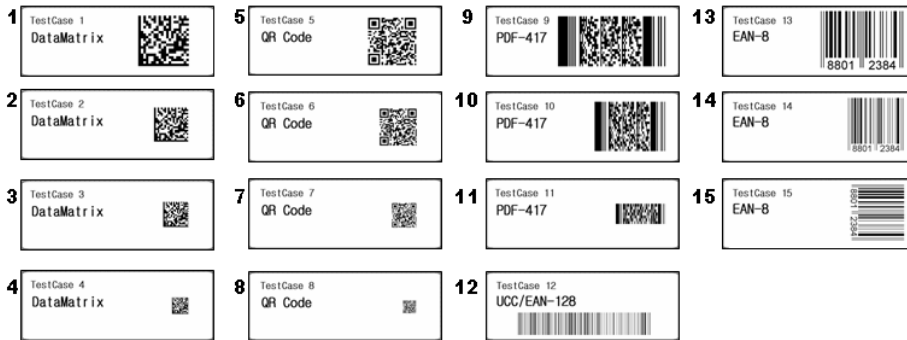
생물자원 보관용기의 최대 표현면적과 바코드 심볼로지의 국제적 범용성을 고려하여EAN-8, UCC/EAN-128, PDF-417, QR Code, Data Matrix를 후보군으로 선정하였고, 선정된 5가지의 바코드 심볼로지에 대한 제안한 데이터 스키마의 적용가능여부를 실험한다.

(그림 9)는 보관용기에 사용할 바코드 심볼로지를 선택하기 위한 실험 데이터와 인식률 실험 과정을 나타낸다. 실험을 위한 데이터는 최대 30자리의 ASCII 데이터를 표현 값으로 하며, 바코드 심볼로지 선정기준에 만족하는 모든 형태의 심볼로지를 테스트하되, 각기 심볼로지의 표준에서 정하는 조정가능 크기 및 밀도 구성에 따르는 형태의 실험 데이터로 15가지가 선별된다. 테스트 케이스 13, 14, 15에 해당하는 EAN-8는 심볼로지 표준 특성에 의해 8자리의 숫자데이터를 사용하고, 나머지 실험 데이터는 공통적으로 30자리를 사용한다. 선별된 실험데이터는 평면에서 인식률과 생물자원 보관용기에 라벨링을 한 상태에서 적용 인식률 실험을 한다. 실험을 위하여 300dpi 해상도의 바코드 전용프린터기와 1, 2차원 바코드 인식이 모두 가능한 바코드 리더를 사용한다. 15가지의 실험 데이터를 20회에 걸쳐 반복 실험을 진행하고, 그 결과를 백분율로 산출하여 (그림 10)에 나타낸다.

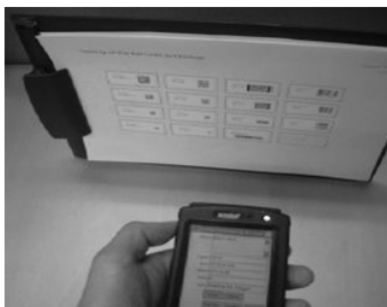
실험결과 높은 수준의 적용상태 인식률을 보이는 테스트 케이스(case#)는 2, 3, 7, 15번이며, Data Matrix, QR Code, EAN-8 바코드 심볼로지가 이에 해당된다. 테스트 케이스 6, 14, 1, 5, 9, 10, 13번은 평면상태에서는 100%의 인식률을 보이고 있지만, 실제 보관용기에 바코드를 라벨링하게 되면 용기 표면의 휘도에 영향을 받아 거의 읽히지 않아 부적합함을 보이고 있다. 그리고 테스트 케이스 4, 8, 11, 12번은 표현 면적 내에 심볼로지가 들어갈 수 있지만 과도한 데이터 밀도를 가지고 있어 인식되지 않는 문제가 있다.

실험결과는 보관용기의 최대 표현면적과 국제적 범용성을

(a) 실험 데이터



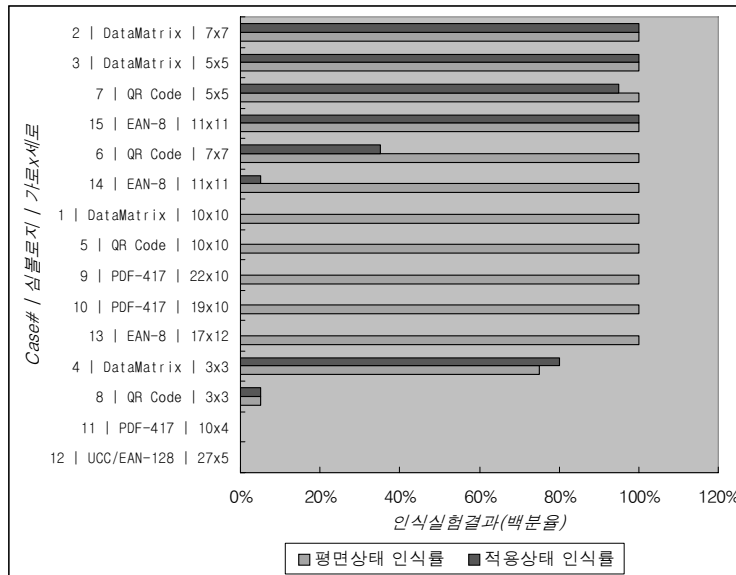
(b) 평면상태 인식을 실험



(c) 보관용기 적용상태 인식을 실험



(그림 9) 바코드 심볼로지 실험 데이터와 인식을 실험



(그림 10) 바코드 심볼로지 테스트 케이스별 인식을 실험결과

고려했을 때의 결과로 Data Matrix, QR Code, EAN-8 바코드 심볼로지가 보관용기 적용상태에서 인식이 뛰어남을 보인다. 그러나 EAN-8은 바코드 심볼로지의 표현 및 인식이 가능할 지라도 생물자원 관리를 위해 사용하기에는 부적합하다고 볼 수 있다. EAN-8 표준에 따르면 8자리 중에서 보존 관리하는 기관이 생물자원을 식별하기 위해 부여할 수 있는 자리는 1자리이므로, 결과적으로 구분 가능한 자원의

수는 10개로 한정되기 때문이다. 따라서 생물자원 관리를 위한 바코드 심볼로지는 Data Matrix, QR Code가 가장 적합하다.

따라서 생물자원의 데이터 요구사항에 부합되며, 자원 관리 환경에 적합한 데이터 표준으로는 본 논문에서 제안한 UCC/EAN-128기반 BRC 데이터 스키마가 적합하다는 것을 알 수 있다. UCC/EAN-128은 기본적으로 1차원 바코드 형

태를 사용하게 되어있으나, 최근 2차원 바코드 중 범용성이 있는 바코드 심볼로지를 사용할 수 있게 표준 제정의 뒤따라 생물자원의 보관용기를 유일하게 식별할 수 있는 태그 인터페이스와 함께 구현이 가능하였다.

7. 결 론

이 논문은 2차원 바코드와 UCC/EAN-128을 이용한 생물자원 자동인식시스템으로 생물자원 컴퓨팅 환경에서의 정확성과 신속성을 높일 수 있음을 보였다. 제안 과정에서 생물자원이 갖는 영하 196℃의 검체 보관 환경과 높은 습도 환경, 실제 도입비용, 그리고 인식해야할 물리적 객체가 태그를 부착하는데 열악한 조건을 모두 충족시킬 수 있는 태그 인터페이스로서 바코드가 가장 적합하며, 바코드 심볼로지는 Data Matrix와 QR Code가 적합함을 인식률 실험을 통해 보였다. 그리고 물리적 객체의 명명 규칙이라고 볼 수 있는 데이터 스키마를 정의하기 위한 요구사항을 기술하고, UCC/EAN-128 기반의 생물자원 데이터 스키마를 정의하였다. 제안된 기술들이 실제 구현에 적용되는지를 보이기 위해 어플리케이션을 개발하고, 실험 및 평가를 수행하였다. 생물자원이 실제 보존되는 영하 196℃, 영하 75℃의 초저온 보존 환경에서 바코드 인식실험을 한 결과 1.6초 내외의 평균 인식률을 보이며, 데이터 스키마는 생물자원 활용 분야의 요구사항을 만족하는 것으로 평가되었다.

연령이나 성별 등 공통의 속성을 가진 인구집단에 대한 대규모의 연구사업이 진행되고, 생물자원을 활용한 연구 활동이 증가함에 따라 정확한 정보전달이 더욱 중요하게 부각되고 있는 현 시점에서, 이 연구는 생물자원과 연관된 다양한 정보전달 과정을 개선시켜 정확성과 신속성을 높여 줄 것이라 기대한다. 향후 과제로는 제시된 생물자원 자동인식 기술을 활용한 생물자원시스템 연계 모델을 설계·구현하여 기업간 생물자원 및 정보의 상호이동성을 높이는 노력이 요구되고 있다.

참 고 문 헌

[1] Kambayashi. Y., Tarumi. H. and Morishita. K., "Digital tags: data with restricted accessibility for e-commerce applications," Information Technology for Virtual Enterprises, Vol.23, No.6, pp.61-70, 2001.

[2] Ricki Lewis, 'Human Genetics: Concepts and Applications,' 7th Ed., McGraw-Hill College, 2007.

[3] 'Biological Resource Centres Underpinning the future of Life Sciences and Biotechnology,' OECD Science & Information Technology, No.7, pp.1-68, 2001.

[4] 'GS1 General Specifications,' Ver.7.0, GS1, 2006.

[5] 'ISBT128 Standard Technical Specification,' Ver.3.1.0, International Council for Commonality in Blood Banking Automation, 2007.

[6] 'The Health Industry Bar Code Provider Applications Standard,' Ver.1.2, Health Industry Business Communications Council, 2006.

[7] 'Information technology - Automatic identification and data capture techniques - Code 128 bar code symbology specification,' ISO/IEC 15417:2007, International Organization for Standardization, 2007.

[8] Tom Werthwine, "HUG work team: Standards Implementation," GS1 Healthcare User Group Conference, No.5, pp.18-19, 2005.

[9] Paul Hebert, Cywinska. A., Ball. S.L. and Dewaard. J.R., "Biological identifications through DNA barcodes," Philosophical Transactions of the Royal Society of London, Vol.B, No. 270, pp.313-321, 2003.

[10] Paul Hebert and Gregory. T.R., "The promise of DNA barcoding for taxonomy," Systematic Biology, No.54, pp. 852-859, 2005.

[11] Michael J. Sailor and Jamie R. Link, "Smart dust: nanostructured devices in a grain of sand," Chemical Communications, Vol.11, pp.1375-1383, 2005.

[12] 'Information technology - Automatic identification and data capture techniques - PDF417 bar code symbology specification,' ISO/IEC 15438:2006, International Organization for Standardization, 2006.

[13] 'Information technology - Automatic identification and data capture techniques - QR Code 2005 bar code symbology specification,' ISO/IEC 18004:2006, International Organization for Standardization, 2006.

[14] 'Information technology - Automatic identification and data capture techniques - Data Matrix bar code symbology specification,' ISO/IEC 16022:2006, International Organization for Standardization, 2006.



주 민 석

e-mail : msc202@cdc.go.kr

2001년 전주대학교 전자계산학(학사)

2006년 충북대학교 전자계산학(석사)

2001년~2004년 한국전자통신연구원
연구원

2005년~현 재 질병관리본부 국립보건연구원
기술연구원

관심분야: 데이터베이스, USN, Biomedical Resources Network,
바이오인포매틱스



류 근 호

e-mail : khryu@dbl-lab.chungbuk.ac.kr
1976년 숭실대학교 전산학과(이학사)
1980년 연세대학교 공학대학원
컴퓨터공학전공(공학석사)
1988년 연세대학교 대학원 전산전공
(공학박사)

1976~1986년 육군군수 지원사 전산실(ROTC 장교),
한국전자통신연구원(연구원), 한국방송통신대학교
전산학과(조교수)
1989년~1991년 Univ. of Arizona Research Staff (TempIS
연구원, Temporal DB)
1986년~현 재 충북대학교 전기전자 및 컴퓨터공학부 교수
관심분야: 시간 데이터베이스, 시공간 데이터베이스, Temporal
GIS, 유비쿼터스 컴퓨팅 및 스트림 데이터 처리,
지식기반 정보검색 시스템, 데이터베이스 보안,
데이터마이닝, 바이오인포메틱스



김 준 우

e-mail : jwkim69@empal.com
1995년 경북대학교 유전공학과(학사)
1997년 경북대학교 대학원 유전공학과
(이학석사)
2002년 고려대학교 생명공학원(이학박사)
2002년~현 재 질병관리본부 국립보건연구원
보건연구사

관심분야: 생물자원 수집 및 관리, chip을 이용한 유전자 발현,
자원 정도관리



김 흥 태

e-mail : bioeel@nih.go.kr
1991년 2월 동아대학교 농과대학 농생물학과
(학사)
1995년 8월 부산수산대학교대학원 생물공학과
(이학석사)
1996년 3월 히로시마대학 대학원 생물권과학연구과
(이학박사)

2002년~현 재 질병관리본부 국립보건연구원 보건연구사
관심분야: 생물자원 수집 및 관리, chip을 이용한 유전자 발현,
자원 정도관리



한 복 기

e-mail : bokghee@nih.go.kr
1981년 부산대학교 생물학과(학사)
1983년 부산대학교 대학원 생물학과(이학석사)
1993년 독일 BONN대학교 세포생물학(이학박사)
1994년~1996년 서울대학교 의과대학
선임연구원

2003년~2004 독일 DSMZ 미생물 및 세포자원연구소
방문연구원
2001년~현 재 고려대학교 의과대학 외래교수
1996년~현 재 질병관리본부 국립보건연구원 보건연구관
관심분야: 유전체, 생물학, 생물자원